



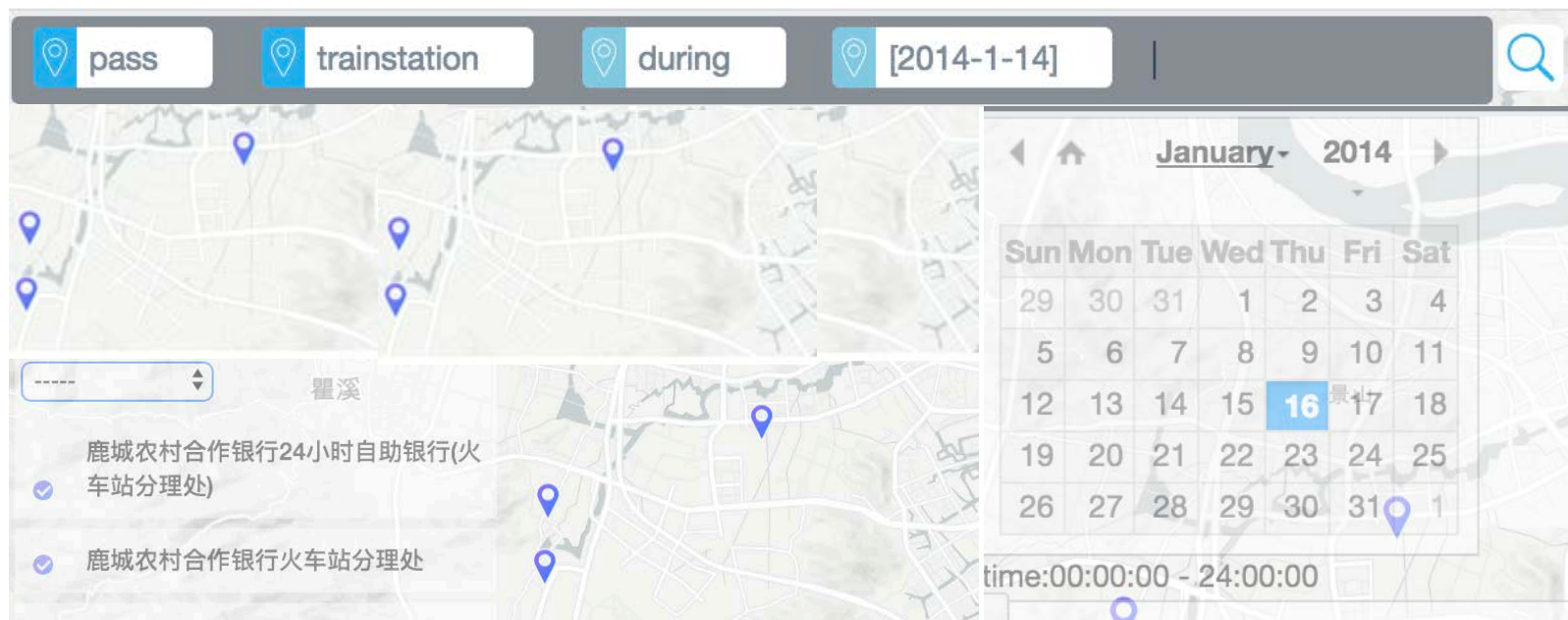
浙江大学  
ZHEJIANG UNIVERSITY

# An Efficient Textualization-based Visual Query Approach For Inaccurate Mobile Trajectories



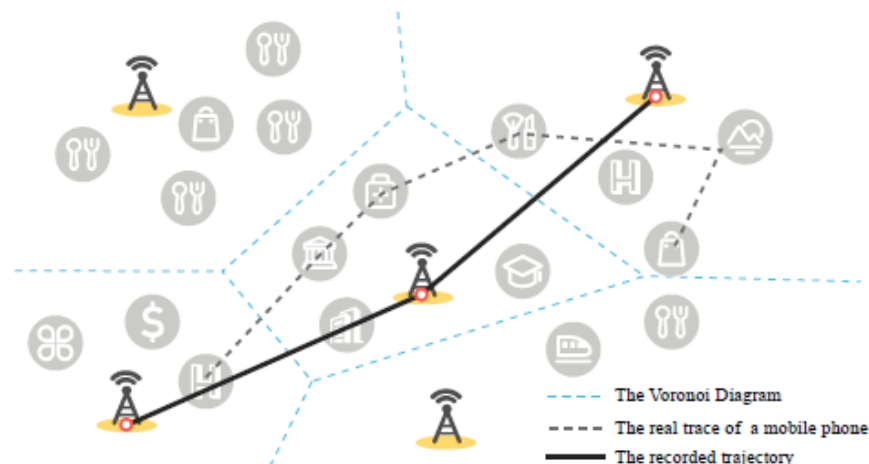
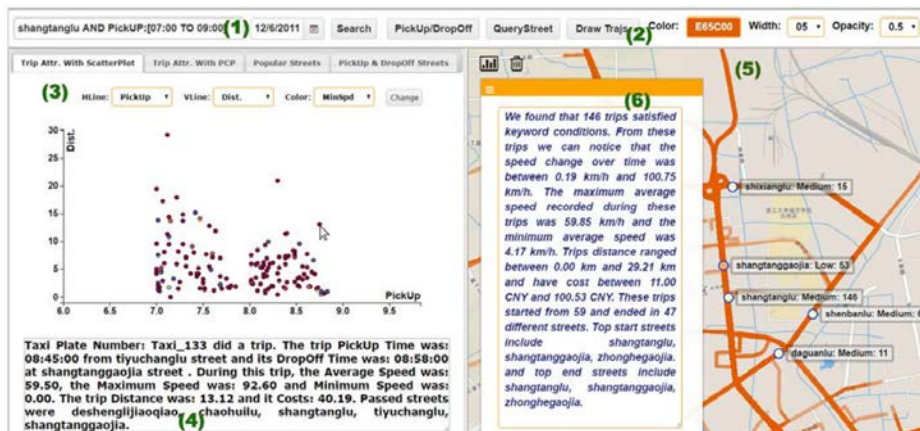
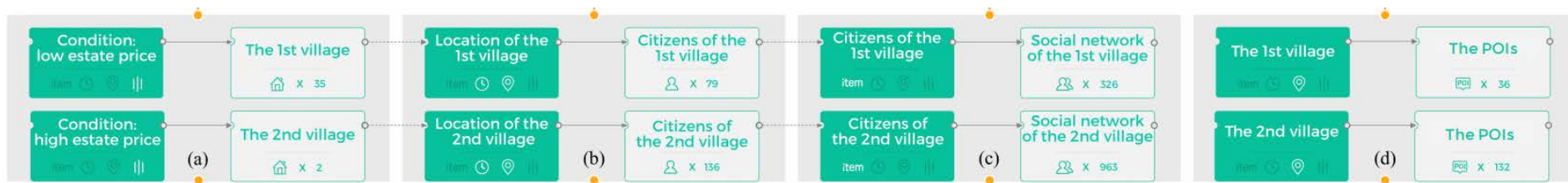
# 背景介绍

- 轨迹数据分析应用中，查询过滤是共有的技术之一。



# 背景介绍

- 轨迹数据分析应用中，查询过滤是共有的技术之一。
- 可视分析系统能够帮助用户表达查询条件
  - 交互式的输入不同类型的查询的条件（VAUD）
  - 用语句输入条件（Semtra j）



# 动机

- 复杂的查询输入过程
- 现有的方法针对于精确的轨迹数据
- 用户可能没有准确的查询目标
  - 对输入数据和查询结果的多个属性进行可视化有助于用户分析查询结果
  - 并不断明确自己的查询目的。



# 解决的问题

复杂的查询输入过程

现有的方法针对于精确的轨迹数据

用户可能没有准确的查询目标

- 对输入数据和查询结果的多个属性进行可视化有助于用户分析查询结果
- 并不断明确自己的查询目的。

- 使用**语句输入**，查询手机基站数据这种**不精确**的轨迹。
- 帮助用户**理解自己的查询输入**，例如查询学校具体对应到了哪些地点条件，并可以在之后的分析中改变自己的输入语句。
- 能够对**查询结果进行探索**，理解查询结果在时空上和语义上的数据属性。

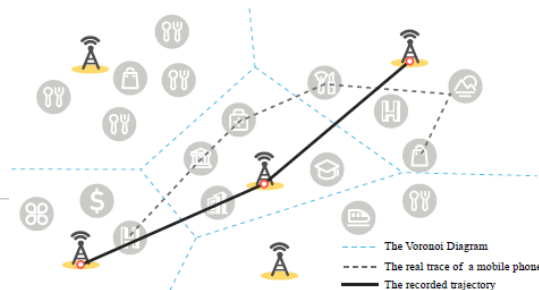


# Contributions

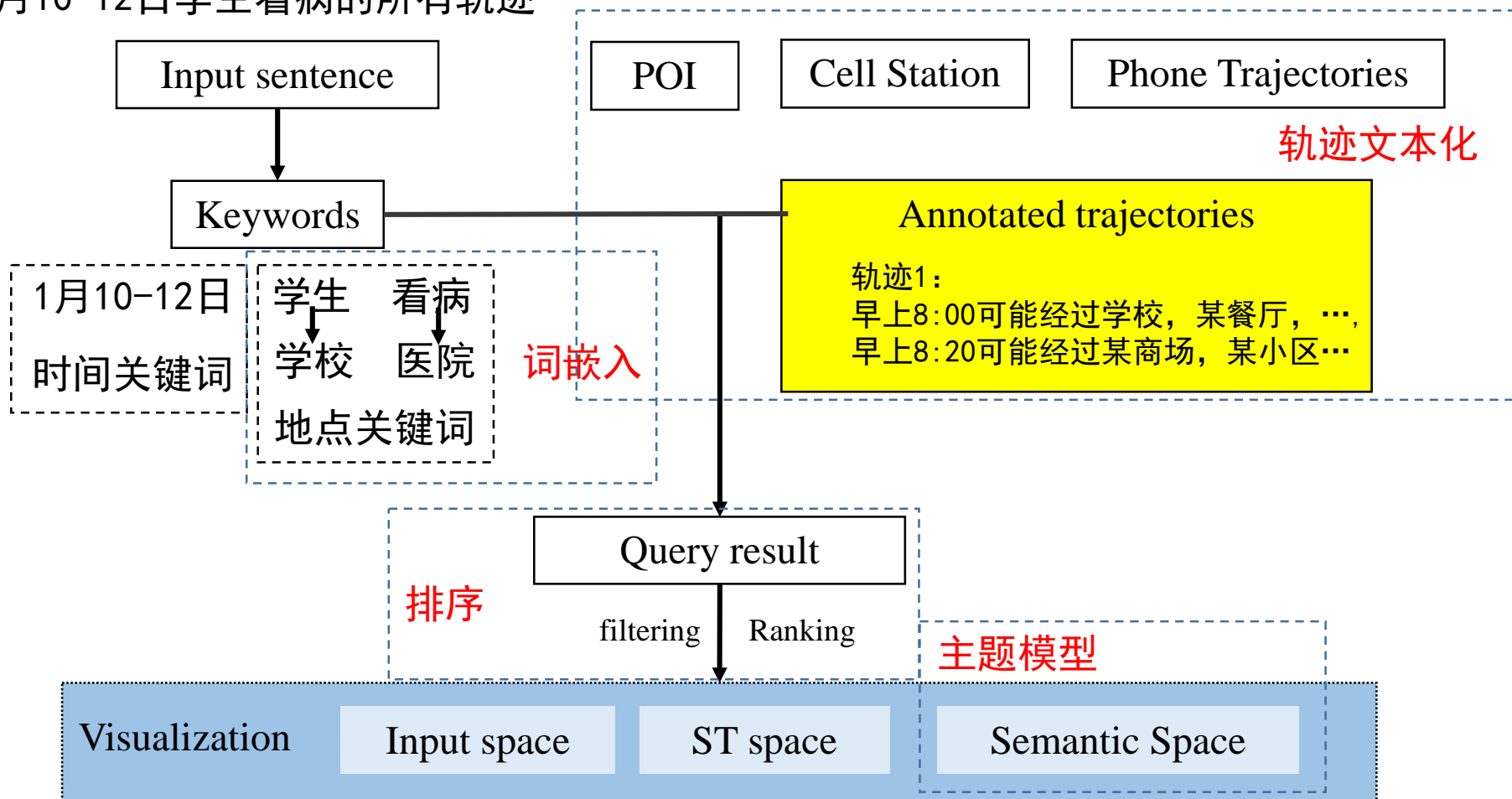
- 一种对不精确轨迹进行 ‘textualization-and-analysis’ 的有效的基于文本输入的查询流程. 并能对查询的结果进行相关性的排序。
- 一个能够进行语义句子输入，查询大规模轨迹的可视分析系统，并对查询结果的时空属性和语义属性进行查看，探索。
- Two use scenarios based on a real-world human mobile trajectory dataset in a mid-size city of China to demonstrate the efficacy of our approach.



# Methods

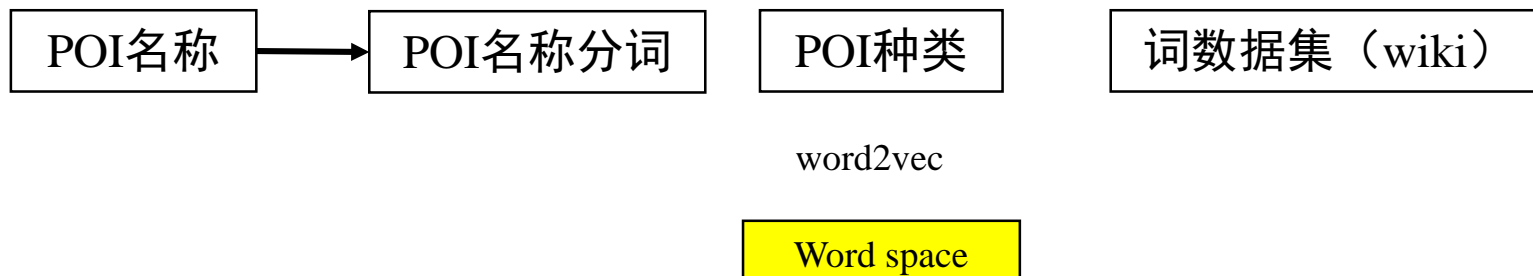


1月10-12日学生看病的所有轨迹



# 关键算法-词嵌入

学生  
↓  
学校  
看病  
↓  
医院



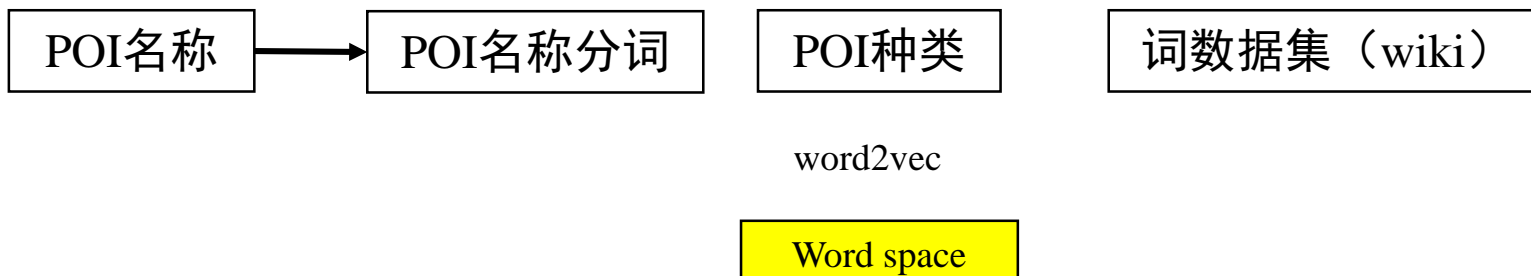
- 预处理部分：将POI种类和POI名称进行word2vec训练嵌入到词空间中。其中POI名称进行分词后训练。
- 对输入的语句，进行分词，然后将不是时间和介词的地点名词与训练好的词向量空间中的词进行比较，查找最接近的K个POI名称或种类的单词。
- 按照时间条件和K个地点单词，查找轨迹。



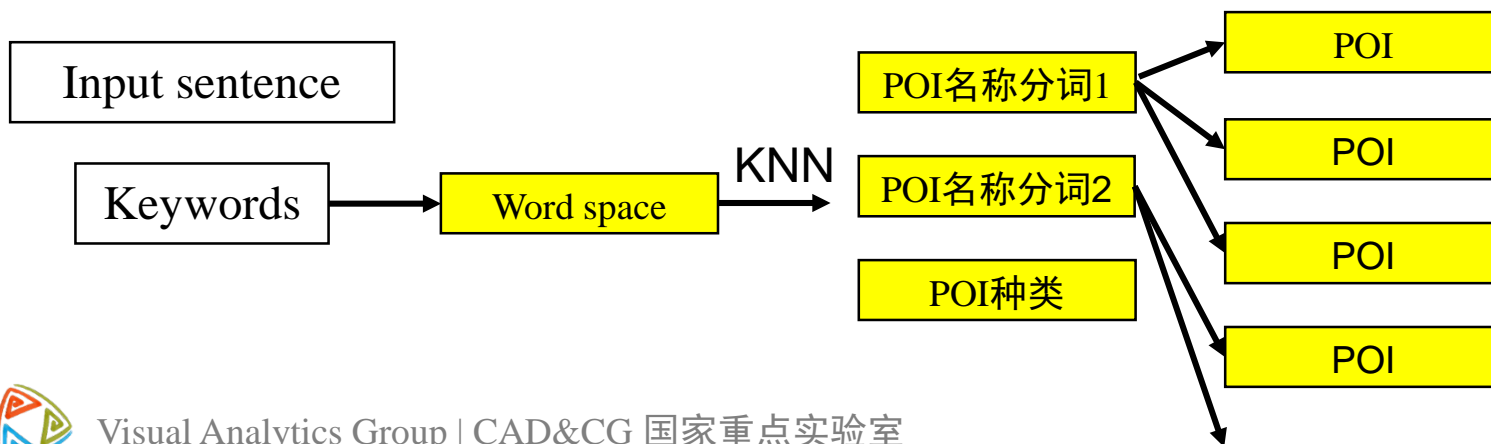


# 关键算法-词嵌入

学生  
↓  
学校  
看病  
↓  
医院

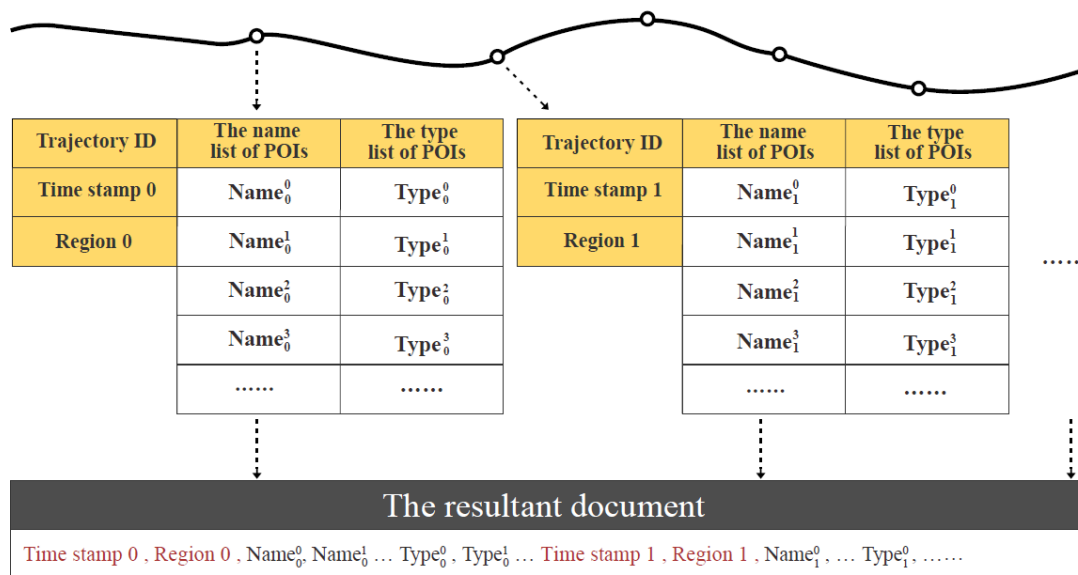


- 预处理部分：将POI种类和POI名称进行word2vec训练嵌入到词空间中。其中POI名称进行分词后训练。
- 查找最接近的K个POI名称或种类的单词。
- 按照时间条件和K个地点单词，查找轨迹。



# 关键算法-轨迹的文本化-数据索引

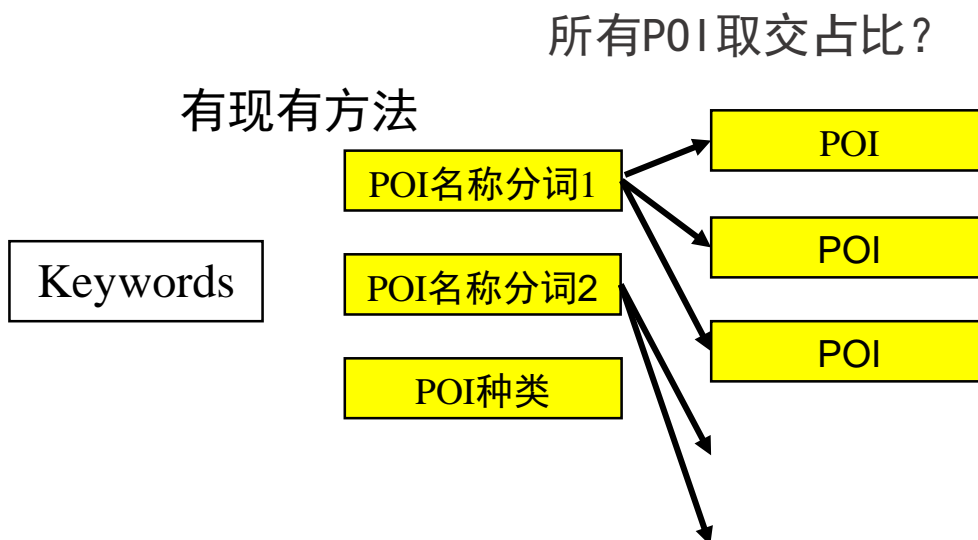
加快以文本的POI名称，时间为查询条件的查询速度  
对POI类型，POI名称，时间空间做轨迹索引



# 关键算法-查询结果排序算法

除过滤掉的轨迹之外：

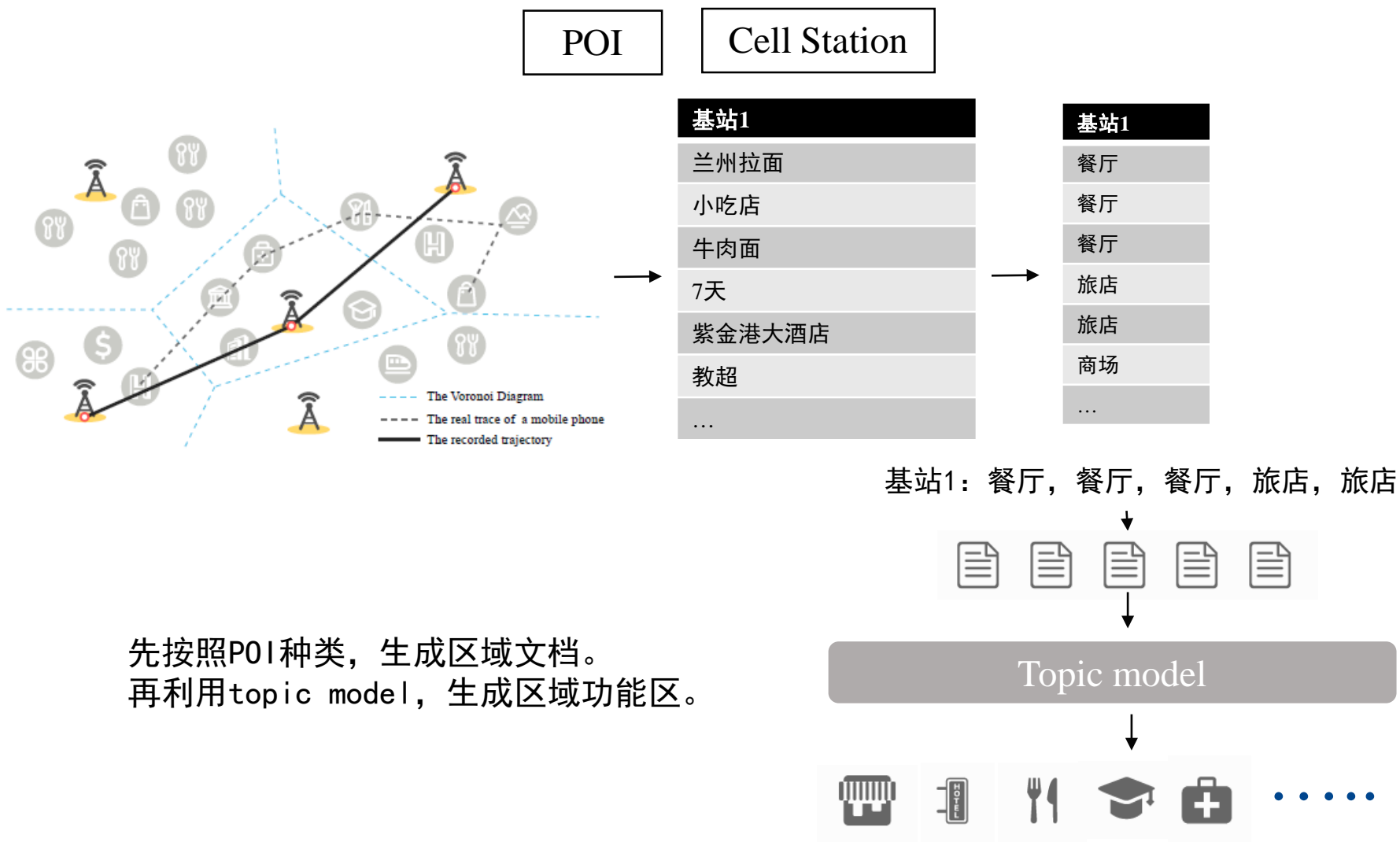
- 与查询条件的差：
  - 词向量空间的距离
  - 分词和POI的距离
- 用户定义的标准轨迹
  - 语义轨迹距离
  - 时间距离
  - 空间距离



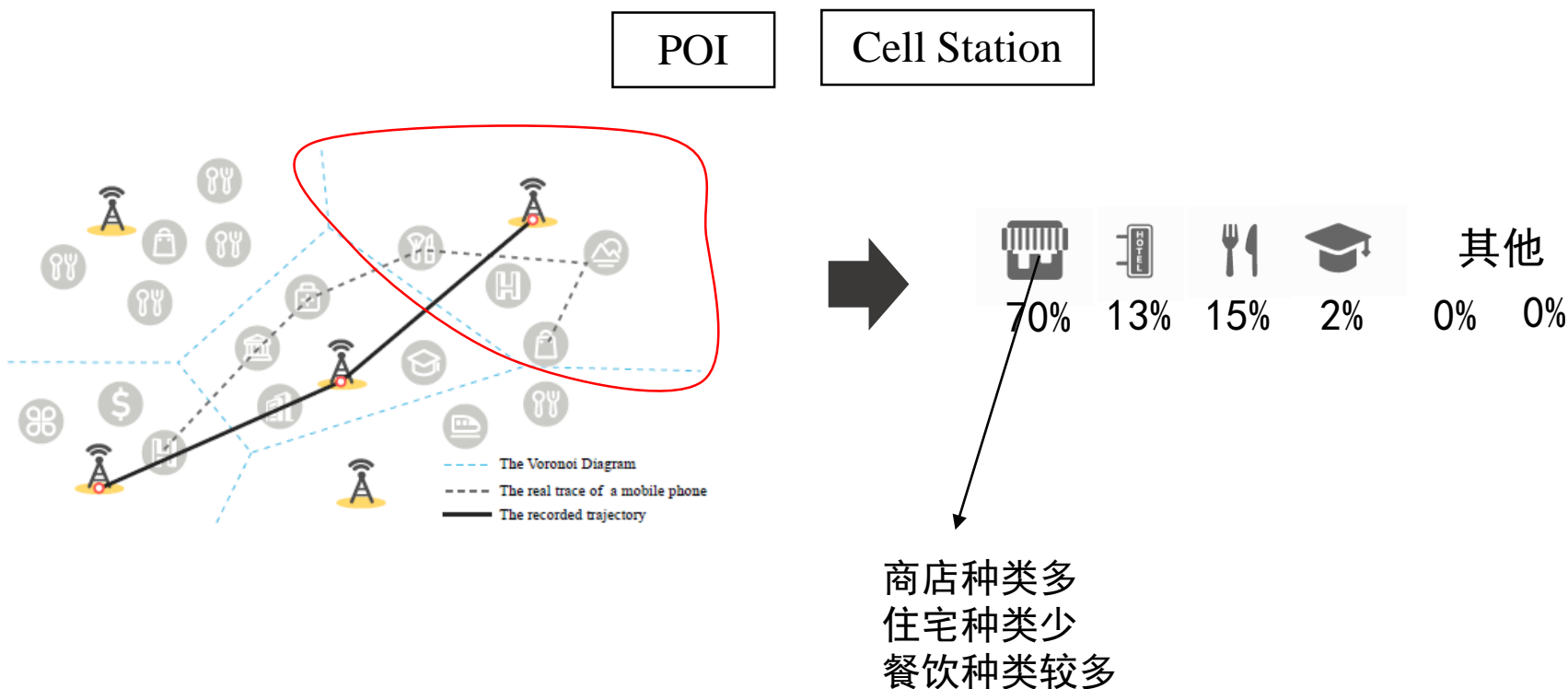
加权方法结合两个部分，用户可调节权重



# 关键算法-主题模型提取区域功能



# 关键算法-主题模型提取区域功能



先按照POI种类，生成区域文档。  
再利用topic model，生成区域功能区。



# 需要进行可视化的信息

语义空间和物理空间两方面

- 理解自己的输入（输入的词向量空间）
- 查询到的轨迹列表，排序
- 轨迹的大致分布（空间，时间，语义）
- 过滤轨迹（时间，空间，语义）
- 选择一部分轨迹看详细的（语义，时间，轨迹）
- 根据要求进行重排序

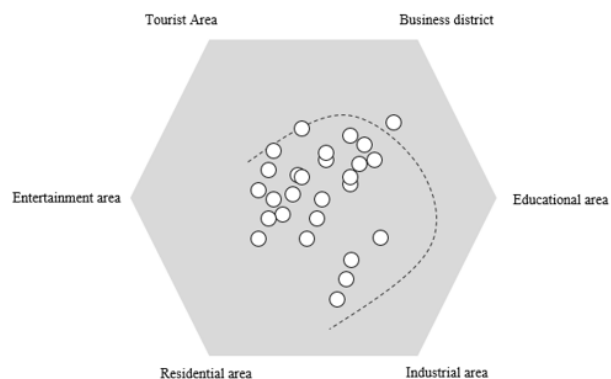


# 可视化设计

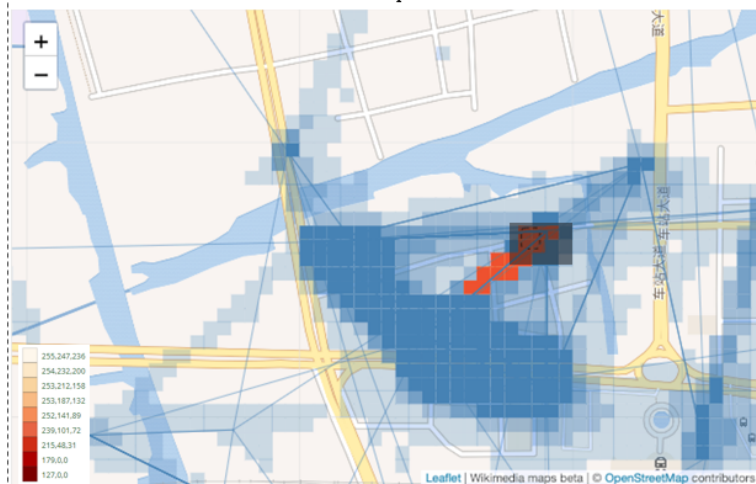
STVA

pass Train station and ZJU on 2014-01-14 morning

Semantic View

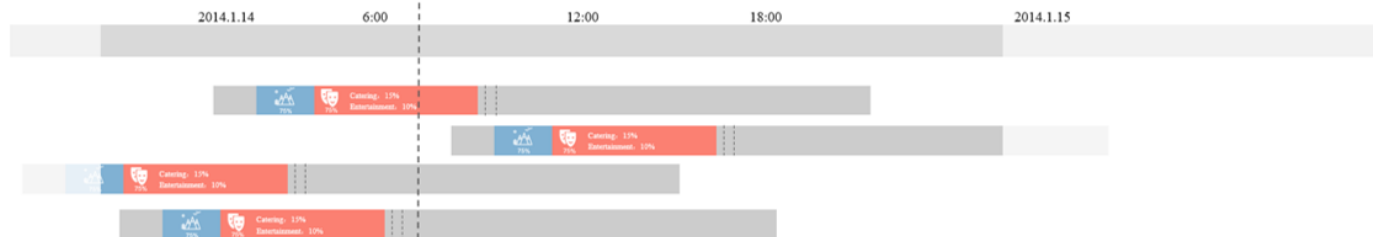


Map View



Functions Categories: Entertainment Area Residential Area Business District Educational Area Industrial Area Tourist/Traffic Area

Semantic Timeline View

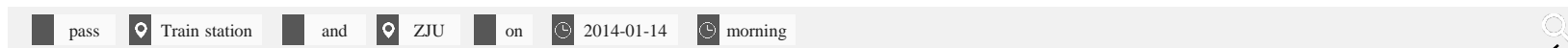


Result List

Listed on Timeline	ID	Correlation	filtered
<input checked="" type="checkbox"/>	485458	98%	
<input type="checkbox"/>	485458	97%	
<input checked="" type="checkbox"/>	485458	96%	
<input type="checkbox"/>	485458	93%	
<input type="checkbox"/>	485458	90%	
<input type="checkbox"/>	485458	90%	
<input checked="" type="checkbox"/>	485458	90%	
<input checked="" type="checkbox"/>	485458	86%	
<input type="checkbox"/>	485458	86%	
<input type="checkbox"/>	485458	86%	
<input type="checkbox"/>	485458	86%	
<input checked="" type="checkbox"/>	485458	86%	
<input type="checkbox"/>	485458	45%	
<input type="checkbox"/>	485458	40%	
<input type="checkbox"/>	485458	40%	
<input checked="" type="checkbox"/>	485458	86%	
<input type="checkbox"/>	485458	45%	
<input type="checkbox"/>	485458	40%	
<input type="checkbox"/>	485458	40%	
<input type="checkbox"/>	485458	40%	
<input type="checkbox"/>	485458	40%	
<input type="checkbox"/>	485458	40%	



# input



查询按钮

说明：

接收用户的输入，然后使用分词算法，自动将句子拆分成词：

一个分词

特殊标记的图标：  
地点或时间



ZJU

词的内容



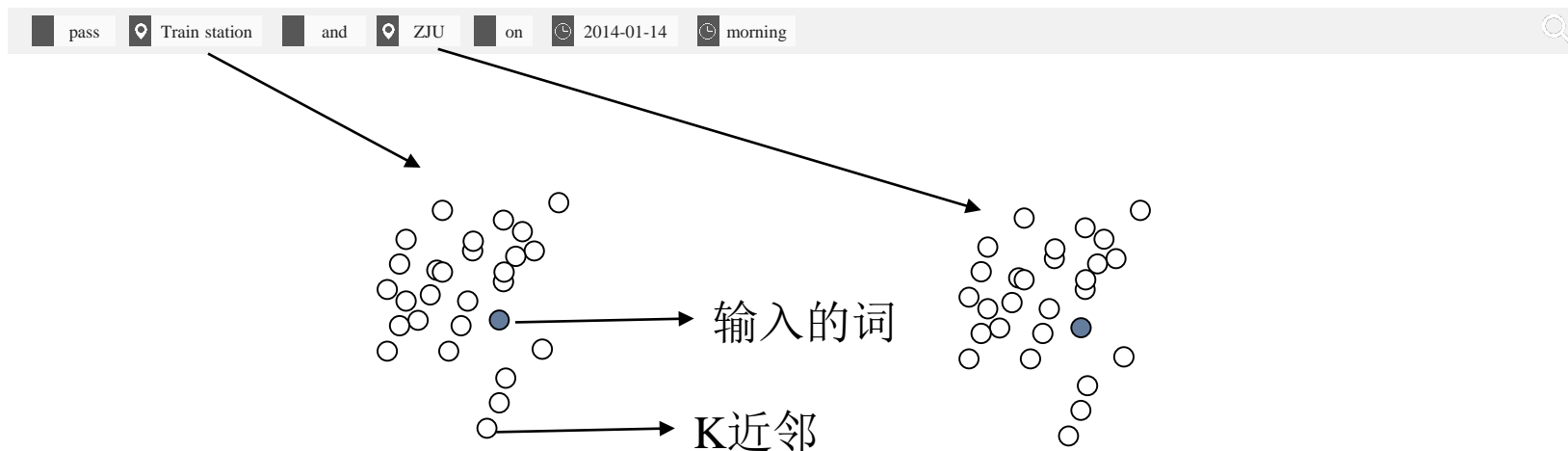
允许的交互：

- 修改词语
- 变更特殊标记
- 点击查询按钮





# Input view-word space



用于理解和调整输入的试图：

- 每一个地点词对应到词空间的K近邻的词。
- T-sne投影
- 可查看  
(可筛选？)

# Result List

查询结果列表：  
根据rank算法计算与输入的相关性。  
然后排序。

加灰的表示在其他界面过滤掉的轨迹  
可以通过点击右上角过滤按钮，选择  
去除这些轨迹或者展示这些轨迹

Checkbox表示是否该轨迹详细信息展  
示在了timeline列表中。

然后就是id和相关性的展示

是否显示过滤

是否展示在  
timeline中

加灰

鼠标hover

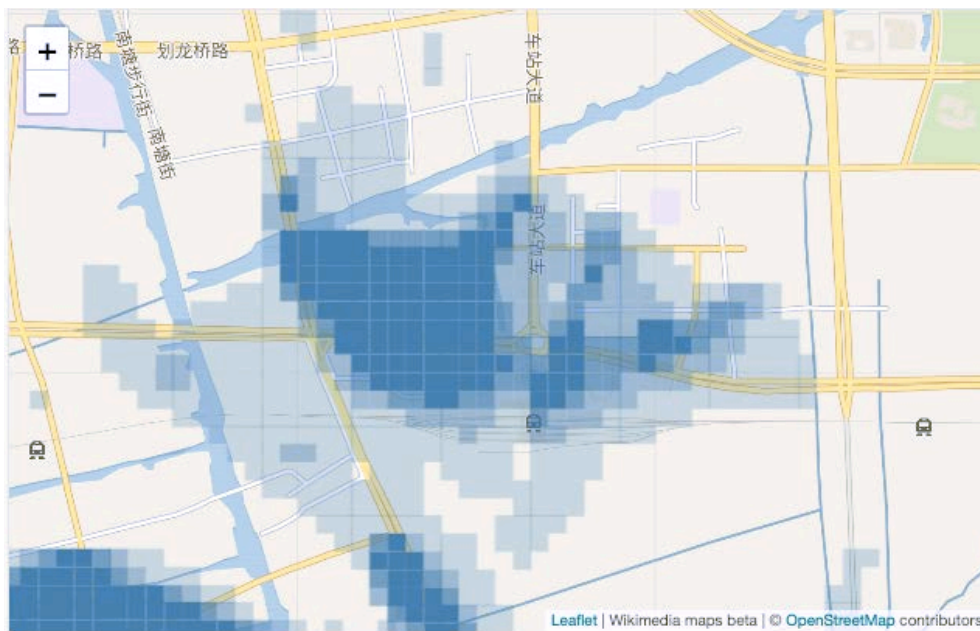
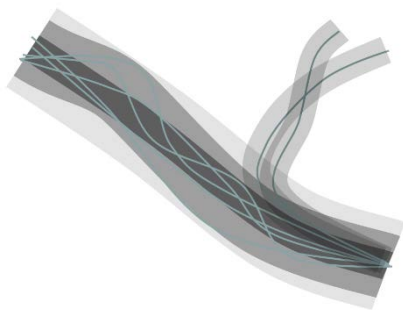
Result List				filtered
Listed on Timeline	ID	Correlation		
<input checked="" type="checkbox"/>	485458	<div></div>		98%
<input type="checkbox"/>	485458	<div></div>		97%
<input checked="" type="checkbox"/>	485458	<div></div>		96%
<input type="checkbox"/>	485458	<div></div>		93%
<input type="checkbox"/>	485458	<div></div>		90%
<input type="checkbox"/>	485458	<div></div>		90%
<input type="checkbox"/>	485458	<div></div>		90%
<input checked="" type="checkbox"/>	485458	<div></div>		86%
<input type="checkbox"/>	485458	<div></div>		86%
<input type="checkbox"/>	485458	<div></div>		86%
<input type="checkbox"/>	485458	<div></div>		86%
<input checked="" type="checkbox"/>	485458	<div></div>		86%
<input type="checkbox"/>	485458	<div></div>		45%
<input type="checkbox"/>	485458	<div></div>		40%
<input type="checkbox"/>	485458	<div></div>		40%
<input type="checkbox"/>	485458	<div></div>		40%
<input type="checkbox"/>	485458	<div></div>		40%
<input type="checkbox"/>	485458	<div></div>		40%



# Map view

对地图进行 缩放、拖动 操作。控制栏中的 **Opacity**滑动条 也可对热力图的颜色深浅按比例控制。

原型：



目的是对**杂乱**的轨迹  
**聚合展示**，方便选择  
同时能够在之后使得**单条轨迹**展示更  
**加清晰**

点击控制栏中的 **Select**按钮，可对感兴趣的区域进行**框选**，虚线框覆盖到的区域被染色，同时会根据周围的值扩大成一个形状。如右图。

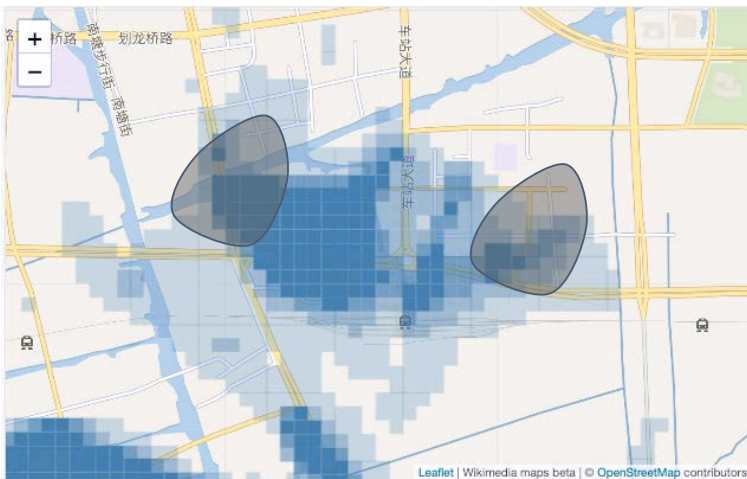
这样的**选择将会过滤出一部分符合要求的轨迹**从而刷新轨迹列表上的排序，没被选中的轨迹将被



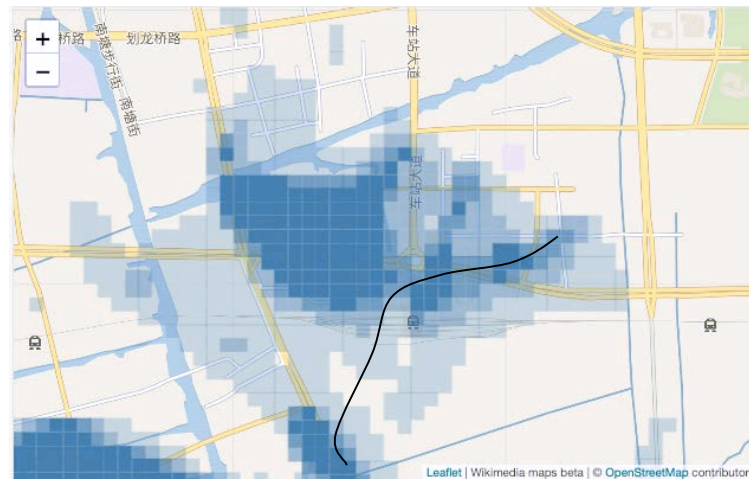
框选

选中的部分轨迹

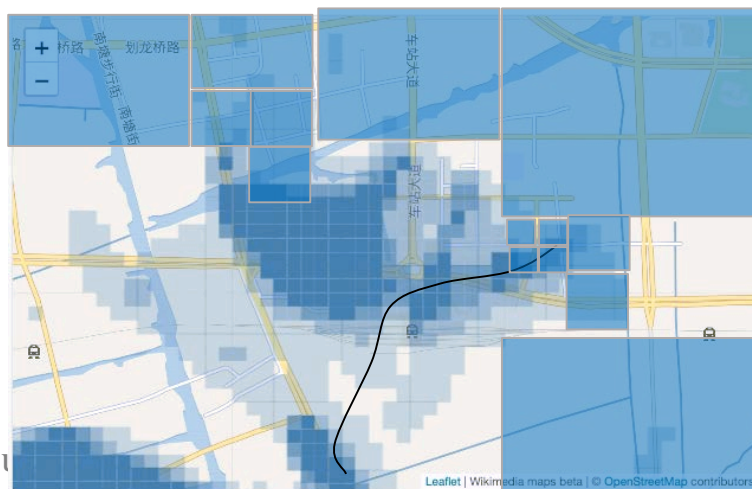
# Map view



分析需求中，要求用户能够对**区域的功能**，**区域内的POI**进行查看。（当鼠标移动到**timeline**的某一段轨迹上时）



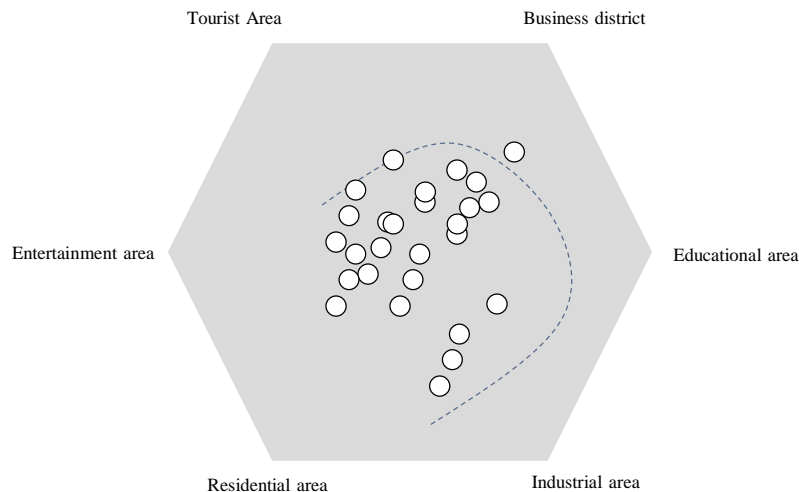
当鼠标移动到某一轨迹或者某一些轨迹时（如有需求），展示轨迹详细内容。（满足时间轴筛选的）



可能的改进算法：使用四叉树计算前面的热力图，**粒度大的地方，更细。**



# Semantic View



轨迹的语义空间：

- 点表示轨迹点。
- 选中的轨迹（timeline里的轨迹）将绘制轨迹线。
- 用户可以对顶点顺序进行拖拽调整。
- 用户可以自己绘制一条标准轨迹，然后所有轨迹按照相似性重新进行排序。刷新列表

•解决方案：投影或力引导

Functions categories



Entertainment area



Residential area



Business district



Educational area



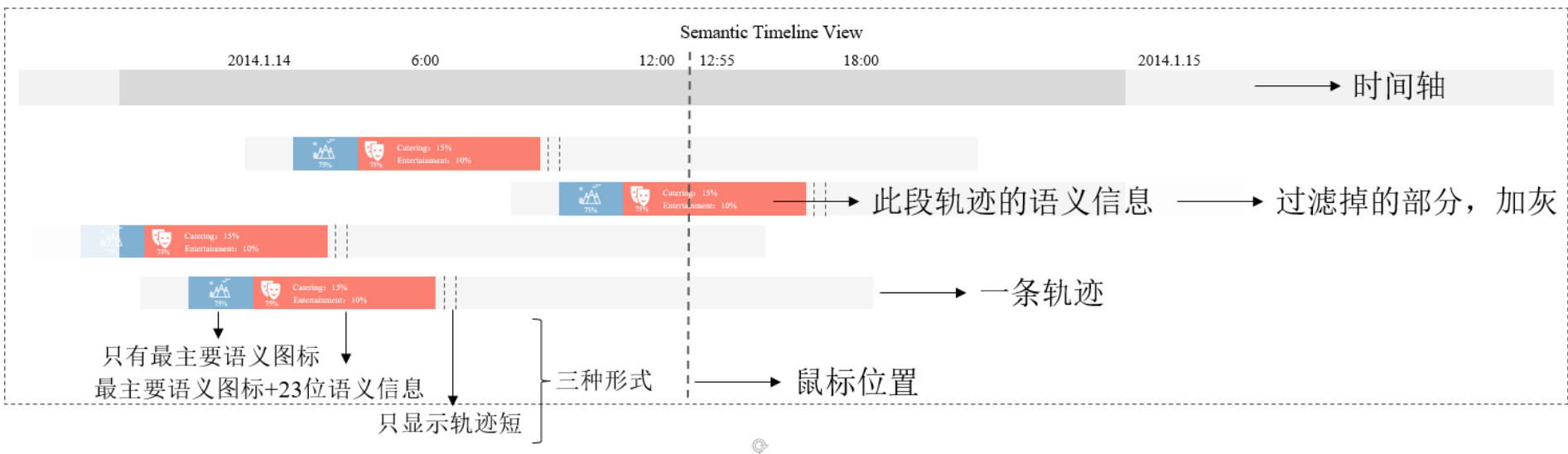
Industrial area



Tourist Area



# Semantic Timeline View



本视图主要功能有2:

- 对单条轨迹的语义进行查看
- 对时间进行选择, 然后过滤

每一个横条表示一条轨迹, 时间轴可放缩, 详细信息按照放缩程度依次展开  
颜色标识功能区, 图例如下

Functions categories

Entertainment area	Residential area	Business district	Educational area	Industrial area	Tourist Area
--------------------	------------------	-------------------	------------------	-----------------	--------------

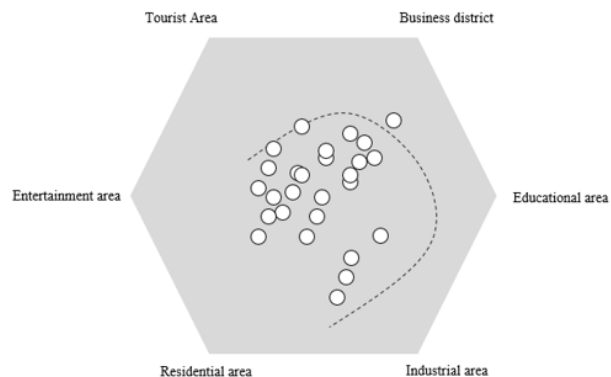


# 可视化设计

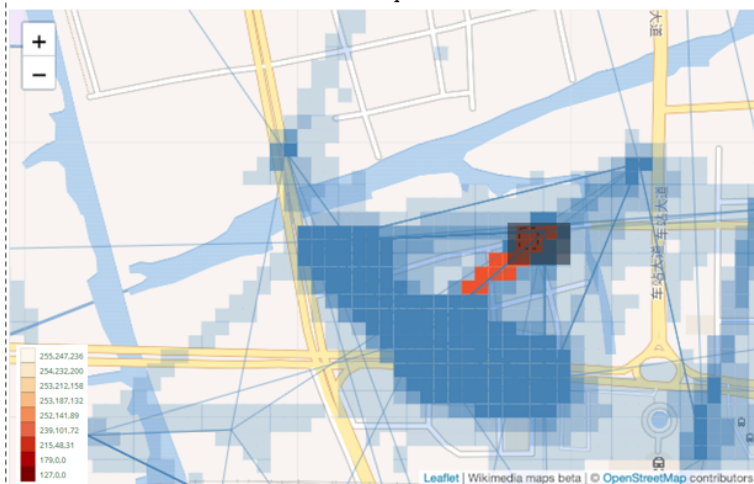
STVA

pass Train station and ZJU on 2014-01-14 morning

Semantic View

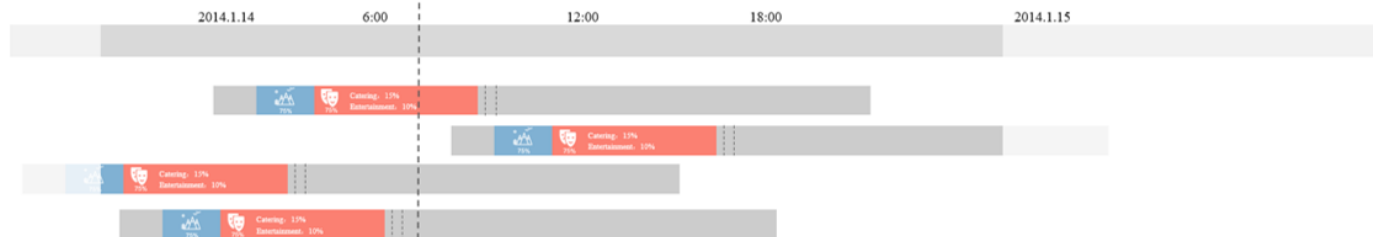


Map View



Functions Categories: Entertainment Area Residential Area Business District Educational Area Industrial Area Tourist/Traffic Area

Semantic Timeline View



Result List

filtered

Listed on Timeline	ID	Correlation
<input checked="" type="checkbox"/>	485458	98%
<input type="checkbox"/>	485458	97%
<input checked="" type="checkbox"/>	485458	96%
<input type="checkbox"/>	485458	93%
<input type="checkbox"/>	485458	90%
<input type="checkbox"/>	485458	90%
<input checked="" type="checkbox"/>	485458	90%
<input checked="" type="checkbox"/>	485458	86%
<input type="checkbox"/>	485458	86%
<input type="checkbox"/>	485458	86%
<input type="checkbox"/>	485458	86%
<input checked="" type="checkbox"/>	485458	86%
<input type="checkbox"/>	485458	45%
<input type="checkbox"/>	485458	40%
<input type="checkbox"/>	485458	40%
<input checked="" type="checkbox"/>	485458	86%
<input type="checkbox"/>	485458	45%
<input type="checkbox"/>	485458	40%
<input type="checkbox"/>	485458	40%
<input type="checkbox"/>	485458	40%
<input type="checkbox"/>	485458	40%
<input type="checkbox"/>	485458	40%



# Discussions

- Mobile phone trajectories
- Python, JS, D3,
- figure: 2-3 cases, interface, background
- Case:
  - 有明确目标的分析任务 去景区的人的轨迹
  - 无明确目标的渐进式查询任务 学生轨迹
- Evaluation: (discussion)
  1. Word2vec dataset
  2. Data size
  3. Time period





# Discussions

## ●方法的局限性

- 轨迹查询结果没有很好的进行聚类。
- 相关性的计算方法需要优化
- KNN方法通过关键词找POI时的排序算法需要优化（找到一个标准）
- 可能到达的POI按距离添加到到达的可能性

## ●可扩展性和未来工作？

- 流程可扩展到类似的接受自然语句的查询过滤系统中
- 语音系统
- 加入词典的可视化生成系统（例如：我要看美国西海岸的犯罪率）
- 加入地理数据的特有词意

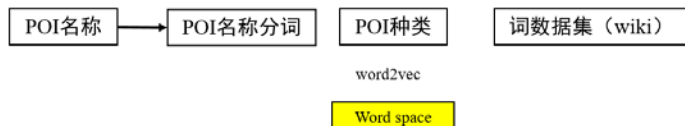
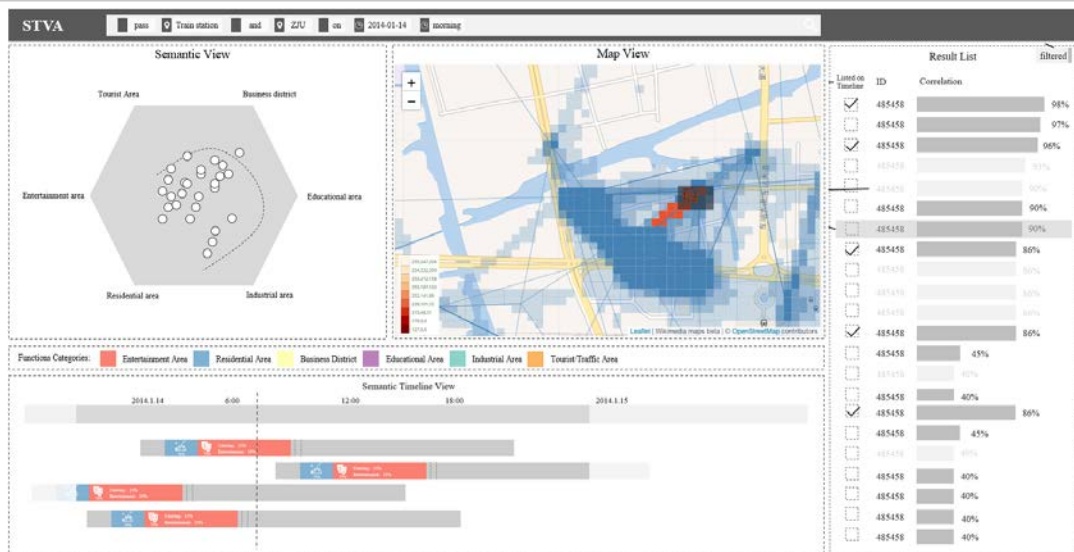


# Remaining Tasks

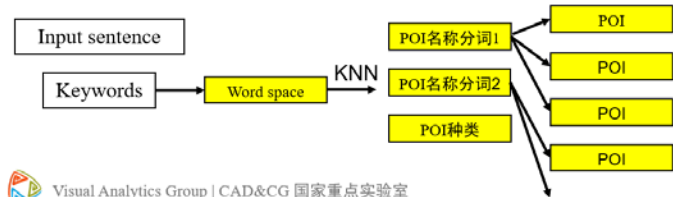
日期节点	需要完成的任务	额外的参与人员（除胜杰，科杰外）
1.22	文章的前两章节完成，系统合并	
1.22	文章的后端方法完成，系统原型完成	
2.18	文章完善和迭代讨论	陈老师，赵老师讨论
3.1	文章前端完成，并完善系统到找case阶段	徐卫霞参与，发送给赵老师整体文章修改。
3.14	找到case并完成casestudy	徐卫霞
3.21	与赵老师讨论修改文章，并制作系统录像	赵老师
3.31	最后一个星期，完善各种可以完善的细节	



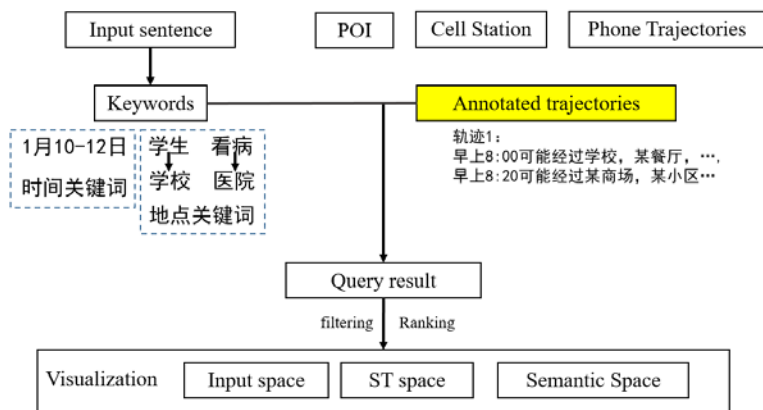
# 请大家给与意见



- 预处理部分：将POI种类和POI名称进行word2vec训练嵌入到词空间中。其中POI名称进行分词后训练。
- 查找最接近的K个POI名称或种类的单词。
- 按照时间条件和K个地点单词，查找轨迹。

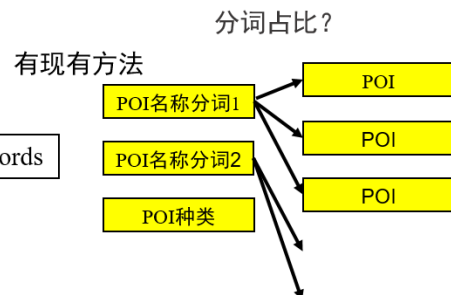


1月10-12日学生看病的所有轨迹



除过滤掉的轨迹之外：

- 与查询条件的差：
  - 词向量空间的距离
  - 分词和POI的距离
- 用户定义的标准轨迹
  - 语义轨迹距离
  - 时间距离
  - 空间距离



加权方法结合两个部分，用户可调节权重

